Projektarbeit

MNI - M@th Desktop Statistik-Projekt

Harald Urwalek und Wolfang Kiegerl, 5bk, 18. März 2005

1

Lineare Regression, Normalverteilung
Mathematische Grundlagen
Open / Close Print

Vergleich der Ping-Daten des Spiels „Natural Selection“
Von der Idee bis zum Diagramm

Die Idee, die Verbindungsdaten - den Ping - von einem Onlinecomputerspiel für unser Mathematik-Projekt zu vergleichen und mit erhaltenen Daten zu arbeiten war relativ schnell geboren - sind wir beide doch begeisterte Computerspieler in unserer Freizeit.

Was genau ist der Ping?

Bei Online Computerspielen ist es notwendig eine möglichst schnelle Verbindung zum verbundenen Server und zu den anderen Mitspielern zu haben. Der Ping ist ein Parameter der die Verbindungsgeschwindigkeit unter den verschiedenen Mitspielern beziehungsweise zu den Servern angibt. Als Regel gilt, je niedriger der Ping Wert - desto besser die Verbindung.
Der Computer schickt kleine Datenpakete an den Server, dieser erkennt die Daten und schickt ebenfalls Datenpakete wieder zurück. Je schneller dies stattfindet desto niedriger der Pingwert.

Welches Onlinespiel?

Eines unserer Probleme war es das richtige Onlinespiel auszuwählen, denn diese gibt es mittlerweile ungefähr so viele wie Sand am Meer. - aber Spaß beiseite, wir mussten uns natürlich für eines entscheiden.
Zuerst wollten wir die Ping Werte des Onlinespieles „Counter Strike“ hernehmen.

Wie wir leider schnell feststellen mussten war dieses Spiel zur Eruierung der Daten überhaupt nicht geeignet, da die Pingwerte vom Programm viel zu schnell aktualisiert wurden um sie zu notieren. Gäbe es eine Möglichkeit diese Daten aufzuzeichnen hätten wir in kürzester Zeit eine Vielzahl von Daten.

Da wir die Daten nur händisch notieren können, mussten wir uns also eine Alternative hierzu überlegen. Aufgrund der schnellen Internetverbindung die wir zu Verfügung hatten, war es für uns kein Problem ein Ersatzprogramm aus dem Internet herunter zu laden. Wir entschieden uns also für das hundertfünfzig Megabyte große Spiele „Natural Selection“.

Bei diesem Programm wird der Ping in regelmäßigen Abständen von einer Sekunde aktualisiert. So war es für uns kein Problem die Zahlen zu erfassen.

Zwar war das aufzeichnen von rund 250 Daten eine sehr zeitaufwändige Arbeit, da wenn man auch nur einen Wert übersah man von neu beginnen musste, da sonst das gesamte Ergebnis unrichtig gewesen wäre.

Wir eruierten die Daten auf zwei verschiedenen Servern. Dies tun wir, um eine Vergleichsmöglichkeit zu schaffen. Der erste Server hat seinen Sitz in Deutschland und ist so für uns relativ schnell erreichbar. Der zweite Server hat seinen Sitz in England - was sich natürlich dementsprechend auf den Ping auswirken wird. Der Ping macht sich nämlich mit einigen „Ausreißern", den so genannten Outliers, bemerkbar. Eine starke Schwankung des Pings, den die Verbindung verursacht wird als „lag" bezeichnet. Diese können natürlich das Ergebnis beeinflussen.

Hier ein Screenshot des ersten - leider erfolglosen - Versuches dem Spiel "Counter Strike" die Pingdaten zu entlocken.

Hier ein Screenshot des vom Onlinespiel "Natural Selection". Dieses Spiel lieferte uns die Daten für unser Projekt.

Mathematik der Linearen Regression

Open / Close

1.

Mit Hilfe der linearen Regression lässt sich der Einfluss einer oder mehrerer erklärbarer Variablen X1,…. Xn, (z.B. X1=Alter, X2 = Geschlecht und X3=Rauchen) auf eine stetige Zielvariable Y (z.B. Y = Blutdruck) statistisch untersuchen. Liegt nur eine erklärende Variable X vor, spricht man von der einfachen linearen Regression (engl.: simple linear regression)

2.

("Mittelwert der Produkte minus Produkt der Mittelwerte")
Bei der Methode der linearen Regression nimmt man an, dass zwischen den beiden Werten ein linearer Zusammenhang besteht, das heißt:
y = ax + b + ein zufälliger Fehler
Die Konstanten a und b werden so bestimmt, dass die Summe der Quadrate der Fehler möglichst klein wird (Methode der kleinsten Fehlerquadrate von C.F. Gauß). Anschaulich können wir uns das so vorstellen, dass wir x und y als Koordinaten von Punkten auffassen und in ein Koordinatensystem einzeichnen. Wir suchen dann die Gerade, die diese Punktwolke am besten annähert Aufgabe kann man mit Hilfe der Differentialrechnung lösen und erhält als Gleichung der Regressionsgeraden:
y = ax + b, wobei

$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$
$$b = \frac{1}{n} \sum y_i - a \frac{1}{n} \sum x_i$$

Die zweite Formel ergibt sich daraus, dass die Regressionsgerade durch den "Schwerpunkt" der Punktwolke geht.
Der Korrelationskoeffizient r liefert ein Maß dafür, wie gut die gegebenen Werte durch diese lineare Funktion angenähert werden. Er ist definiert durch

$$r = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n \sum x_i^2 - (\sum x_i)^2)(n \sum y_i^2 - (\sum y_i)^2)}}$$

Der Wert von r liegt immer zwischen -1 und 1. Dabei bedeutet
r nahe bei 1: starke positive Korrelation (je größer x, umso größer y)
r nahe bei -1: starke negative Korrelation (je größer x, umso kleiner y)
r nahe bei 0: schwacher oder gar kein Zusammenhang
Manchmal verwendet man auch das Bestimmtheitsmaß r^2. Es gibt an, welcher Anteil der

Abweichungen vom Mittelwert durch die Korrelation erklärt wird,

Literatur zur Linearen Regression

Open / Close

1.

http://www.thieme-connect.com/ejournals/pdf/dmw/doi/10.1055/s-2002-32818.pdf

(nur Teilauszug)

2.

http://members.chello.at/gut.jutta.gerhard/kurs/statistik2.htm

Mathematik der Normalverteilung

Open / Close

1.

Die Bedeutung der Normalverteilung wurde Ende des 18. Jahrhunderts im Rahmen von Messfehleruntersuchengen, die vor allem von Carl Friedrich GAUSS (1777 - 1855) durchgeführt wurden, bekannt.

Der belgische Astronom und Statistiker Lambert Adolphe Jacques QUETELET (1796 - 1874) stellte fest, dass überraschend viele Zähl- und Messergebnisse Häufigkeitsverteilungen aufweisen, die dieser Verteilung ähnlich sind. QUETELET beschäftigte sich vorwiegend mit der Untersuchung des menschlichen Körperbauchs und ging dabei von einem Durchschnitts- bzw. Normalmenschen aus.

Die Verteilung wird daher Normalverteilung oder GAUSS-Verteilung genannt.

Am einfachsten kann man sich die Entstehung der Normalverteilung als Grenzübrgang aus einer Binomialverteilung am so genannten GALTON_SCHEN Nagelbrett (benannt nach Francis GALTON (1822 - 1911), englischer Reisender und Naturforscher) vorstellen, das in seinen wesentlichen Konstruktionsmerkmalen in Figur 1 (unten) dargestellt ist.

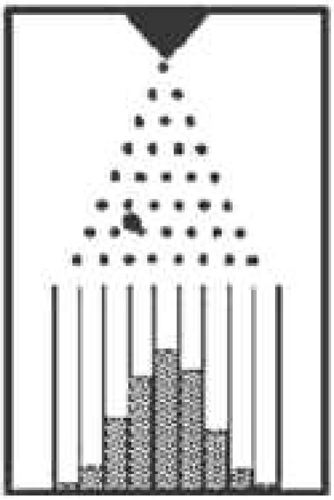
Aus einem Trichter fallen, unabhängig voneinander, gleich große Kugeln durch ein System von Nagelreihen, deren jeweils übereinander liegende Nagelreihen um einen halben waagrechten Nagelabstand verschoben sind. Unter dem System der Nagelreihen befinden sich dicht nebeneinander senkrechte Schächte, deren Breite gleich dem Nagelabstand un ein wenig breiter als der Kugeldurchmesser ist.

Eine Kugel, die im freien Fall den Trichterausgang verlässt, wird bei jeder Nagelreihe mit der Wahrscheinlichkeit p=1/2 nach rechts bzw. nach links abweichen. Nach dem Fall über n Nagelreihen werden die Kugeln die symmetrisch zum Trichterausgang angeordneten Schächte mit den Nummern 0, 1, 2,…, n füllen.

Die Entstehung einer solchen Binomialverteilung kann man sich an einem Gedankenexperiment klarmachen: lässt man 2^6 = 64 Kugeln über ein aus 6 Nagelreihen bestehenden Nagelbrett fallen, so hat man unterhalb dieses Systems in den Schächten mit den Nummer 0, 1, 2, 3, 4, 5, 6 von links nach rechts folgende Kugelzahlen zu erwarten: 1, 6, 15, 20, 15 , 6, 1.

Man kann jede der Nagelreihen als eine der vielen nicht zu beherrschenden Betriebseinflusse deuten, denen ein technischer Prozess, z.B. die Fertigung von Drehteilen auf einem Automaten, ausgesetzt ist.

galtonsches Brett [nach Sir F. Galton], Vorrichtung zur Veranschaulichung der Binomialverteilung: Kugeln werden durch gleichmäßig angeordnete Hindernisse \ (Nägel) in Kammern verteilt. Damit kann man annähernd eine Normalverteilung erreichen.



2. und 3.

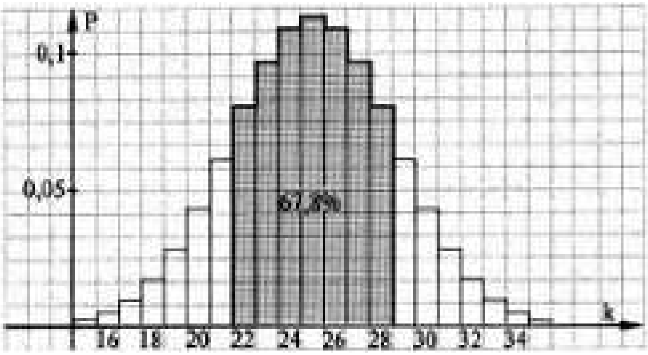
$\mu = n * p$ Mittelwert einer Population

$\sigma = \sqrt{n * p * (1 - p)}$ Standardabweichung einer Population

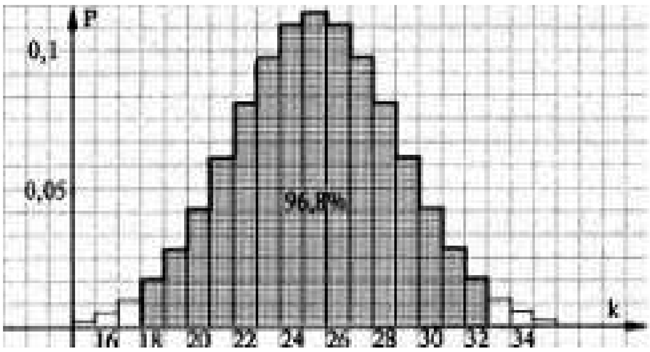
Der Mittelwert μ verschiebt die Verteilung.
σ ist die Abweichung vom Mittelwert. Je kleiner σ, umso höher und spitzer ist die Verteilung. Wenn σ groß ist, ist sie breiter.

Mit dieser Standardabweichung kann man drei so genannte σ – Umgebungen definieren.

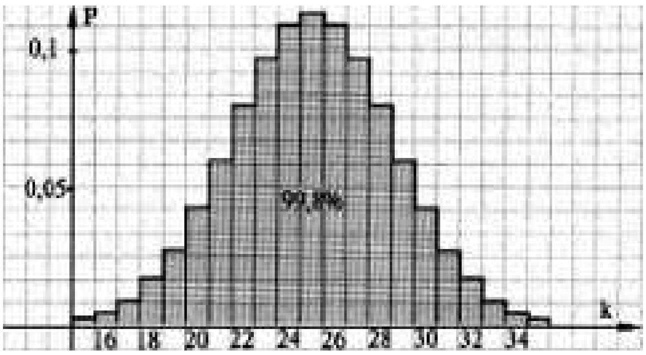
In der 1. σ – Umgebung werden 68 % aller Daten erwartet,
P $(\mu - \sigma = X = \mu + \sigma)$



in der 2. σ-Umgebung 95,5%
P $(\mu\text{-}2\sigma = X = \mu\text{+}2\sigma)$



und in der 3. σ-Umgebung 99,7%.
P $(\mu\text{-}3\sigma = X = \mu\text{+}3\sigma)$



Literatur zur Normalverteilung

Open / Close

1.

Mathematik und ihre Anwendung in der Wirtschaft, Steiner - Weilharter, 2. Auflage 1998, Band 4, Reniets Verlag

2.

Schulheft

3.

http : // www.mathe – aufgaben.de / mathehilfen / mathe – abitur / Stochastik / 34311 %20 Sto %209 %20 BinVert %202 %20 %20 EW %20 SOD.pdf

2 Reallife-Beispiel zur Lineare Regression

Open / Close Print

Aufgabenstellung

Open / Close

Im Sinne unseres Mathematik Projektes scheint für uns nun die Zeit gekommen zu sein, die vielleicht größte und wichtigste Frage im Leben eines Computerspielers zu lösen.

"Wer is Schuld am schlechten Ping?" (Zitat eines anonymen Spielers)

Diese Frage stellt sich ein Computerspieler ungefähr 2 Mal am Tag. Wir wollen nun herausfinden ob für den schlechten Ping wir selbst (beziehungsweise unsere Internetverbindung oder unser PC) verantwortlich sind, oder ob die Ursache ein paar hundert Kilometer weiter weg - nämlich am Server anzutreffen ist.

Um diese Frage zu lösen, bedienen wir uns der linearen Regression. Mit Hilfe dieser, lassen sich die Ping Werte der Spieler unterneinander vergleichen und in Folge dessen haben wir die Möglichkeit Zusammenhänge zu erkennen und zu notieren.

Bei unserem ersten Versuch vergleichen wir unseren eigenen Ping mit dem Ping eines Mitspielers der dem unseren ähnelte.

Bei unserem zweiten Versuchen nehmen wir zusätzlich zu den eigenen Daten den Ping eines Mitspielers der ungefähr um 25 höher ist als der unsere.

Wir werden von beiden Mitspielern 15 Daten notieren.

Haben die Pings Gemeinsamkeiten, so kann man davon ausgehen das der Server eine gewisse Schwankung verursacht und daher für den guten bzw. schlechten Ping verantwortlich ist.

Sollten sich allerdings keine Gemeinsamkeiten feststellen so lässt sich sagen das der Grund für Schwankungen quasi "hausgemacht" ist.

Mathematische Modelierung

Open / Close

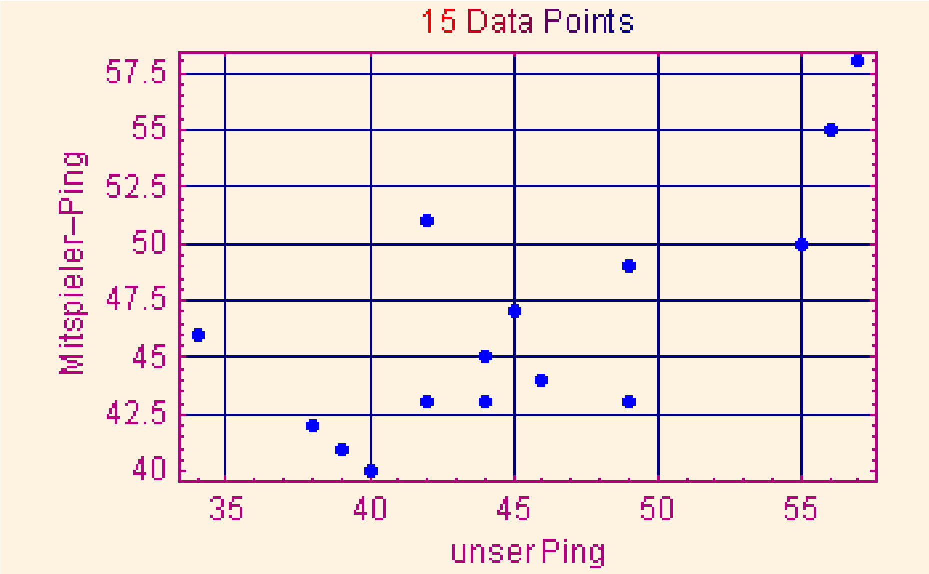
Note: Nachdem wir also wie oben angeführt die Daten eines Mitspielers ermittelten fügten wir diese hier in *Mathematica* ein um mit ihnen arbeiten zu können.

Spieler mit ähnlichem Ping:

Input ▹
```
data = {{46, 44}, {57, 58}, {49, 43}, {34, 46},
  {38, 42}, {42, 43}, {40, 40}, {39, 41}, {55, 50}, {45, 47},
  {49, 49}, {44, 45}, {42, 51}, {56, 55}, {44, 43}}; Add ;
```

Note: Um das ganze graphisch darzustellen, ploten wir dei Daten. Jeder Datenpunkt ist blau eingezeichnet.

Input ▹
```
MDPlotData[data ,
 FrameLabel → {"unser Ping", "Mitspieler-Ping"},
 PlotStyle → {Blue, PointSize[0.02]}];
```



Mit dem freien Auge ist hierbei für uns nicht deutlich zu erkennen, ob die Daten miteinander verwandt sind.

Note: Bei der Regressionsgerade rechnet *Mathematica* für uns den a-Wert und den b-Wert aus. Diese Werte müssen wir berechnen, wenn wir nur den x-Wert kennen und den y-Wert berechnen wollen.

Input ▹
```
Switch to X[y] ;
Clear[x];
regrLineY[x_] = Fit[data, {1, x}, x]
```

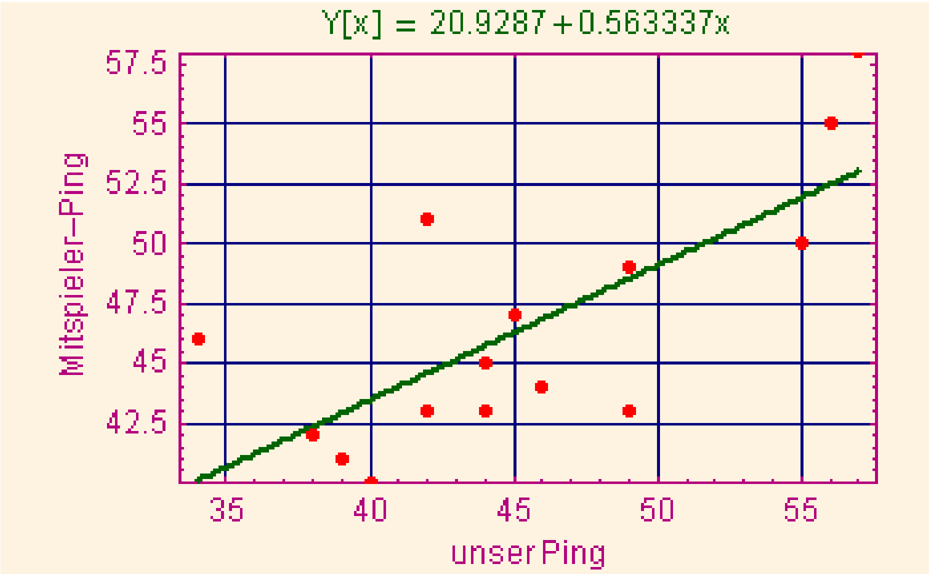
```
20.9287 + 0.563337 x
```

Die Regressionsgerade lautet 20.9287+0.563337 x.

Note: Hier zieht der Computer die Regressionsgerade, indem er die kleinste Summe der Fehlerquadrate berechnet.

```
Both Least Squares Lines ;
Clear[x]; regrLineY[x_] = Fit[data, {1, x}, x];
```

Input ▹
```
MDSPlotDataRegressionLineY[data, {x, regrLineY[x]},
    PointStyle → {Red, PointSize[0.02]},
  FrameLabel → {"unser Ping", "Mitspieler-Ping"},
    PlotStyle → {{DarkGreen, Thickness[0.01]}}];
```



Aufgrund der Regressionsgerade kann man vermuten, dass die Daten einen Zusammenhang haben. Dabei handelt es sich allerdings nur um eine Vermutung. Wenn der r-Wert zwischen 0.6 und 1 ist, besteht ein starker direkter linearer Zusammenhang. In unserem Fall beträgt der r-Wert rund 0.734. Dies haben wir händisch und mit dem Computer berechnet.

Note: Nun nehmen wir an, dass unser Ping 45 beträgt und berechnen den wahrscheinlichen Ping unseres Mitspielers. Danach werden wir diesen errechneten Punkta auf der Regressionsgeraden blau einzeichnen.

```
Switch to X[y] ;
```

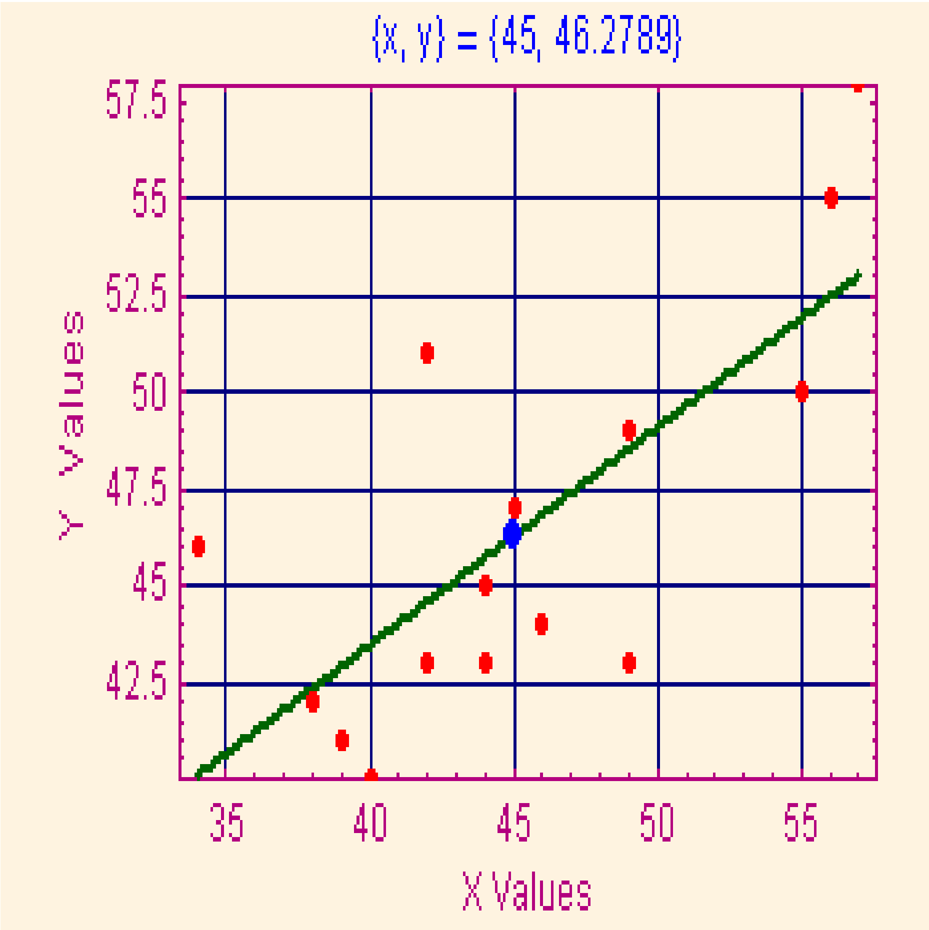
Input ▹
```
Clear[x];
regrLineY[x_] = Fit[data, {1, x}, x];
regrLineY[45]  (* x value MUST be in data x-range *)
```

46.2789

```
Switch to X[y] ;
xvalue = 45; (* xvalue MUST be in data x-range *)
```

Input ▹

```
MDSPlotDataRegressionLineYPt[data, xvalue,
    Epilog → {Blue, PointSize[.03]}];
```



Answer: Unser Ping von 45 entspricht einem Ping unseres Mitspielers von VPI von 46.2789.

Spieler mit durchschnittlich um 25 höheren Ping:

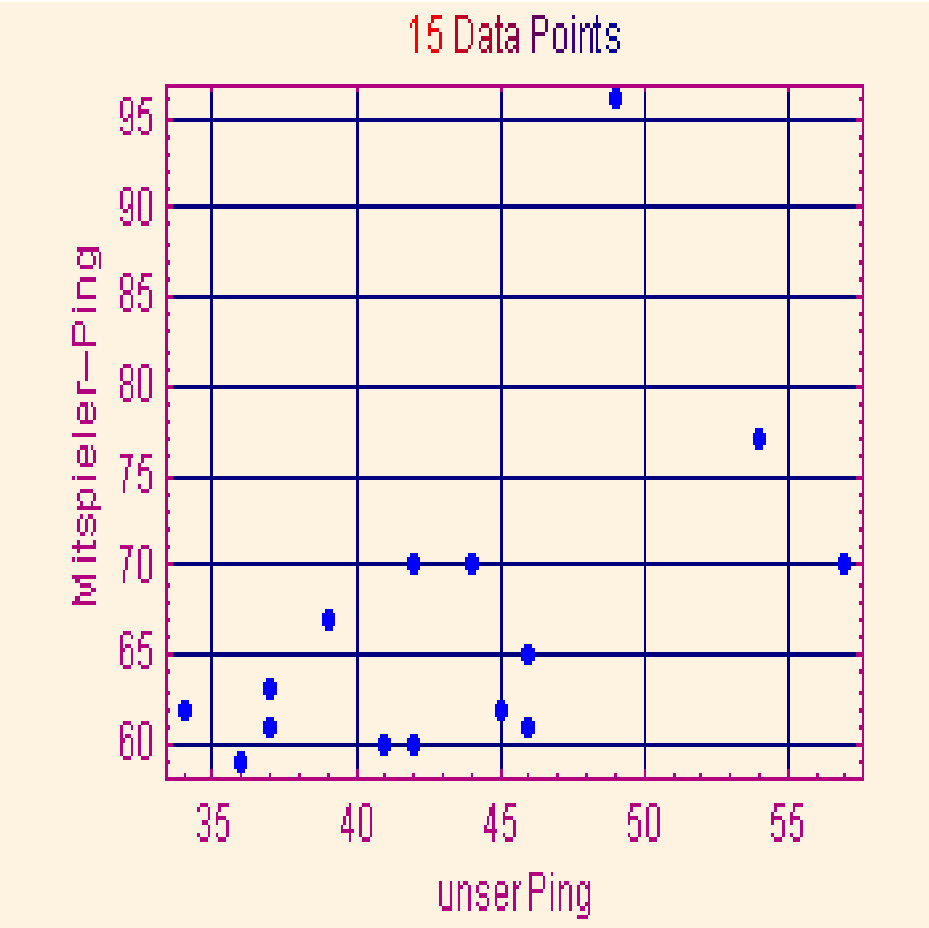
Note: Hier notieren wir die zweite Datenreihe.

Input ▹

```
data = {{44, 70}, {54, 77}, {37, 63}, {39, 67},
  {42, 70}, {46, 61}, {46, 65}, {57, 70}, {34, 62}, {49, 96},
  {42, 60}, {45, 62}, {37, 61}, {41, 60}, {36, 59}}; Add ;
```

Note: Auch hier möchten wir die gesammelten Daten graphisch veranschaulichen, und versuchen anschließend eine Aussage über den Zusammenhang der Daten zu treffen.

Input ▹
```
MDPlotData[data ,
FrameLabel → {"unser Ping", "Mitspieler-Ping"},
 PlotStyle → {Blue, PointSize[0.02]}];
```



Anhand der graphischen Darstellung lässt sich kein Zusammenhang zwischen den Zahlen erkennen. Ein paar Datensätze scheinen linear zu steigen, andere wiederrum weichen stark ab.

Note: Jetzt berechnen wir die Regressiongerade.

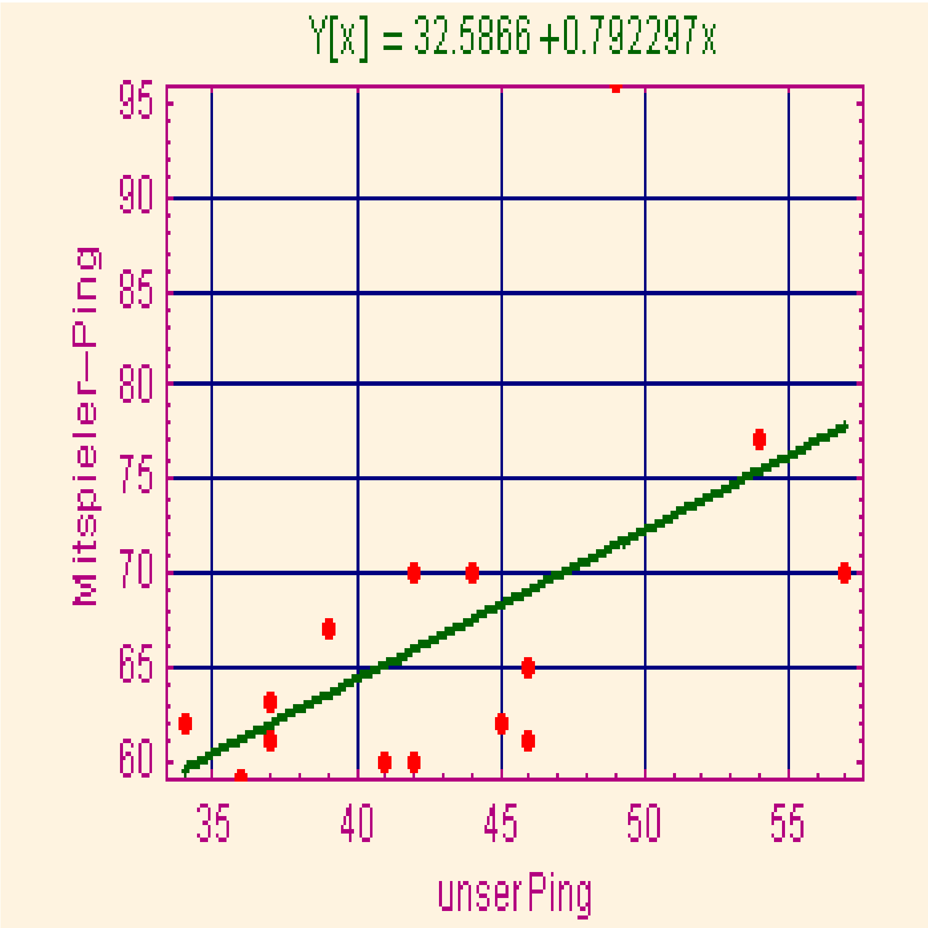
Input ▹
```
Switch to X[y] ;
Clear[x];
regrLineY[x_] = Fit[data, {1, x}, x]
```

```
32.5866 + 0.792297 x
```

Die Regressionsgerade beträgt 32.5866+0.792297 x.

```
Both Least Squares Lines ;
Clear[x]; regrLineY[x_] = Fit[data, {1, x}, x];
```

Input ▹
```
MDSPlotDataRegressionLineY[data, {x, regrLineY[x]},
    PointStyle → {Red, PointSize[0.02]},
  FrameLabel → {"unser Ping", "Mitspieler-Ping"},
    PlotStyle → {{DarkGreen, Thickness[0.01]}}];
```



Zu unserem großen Überraschen errechnen wir händisch und am Computer, dass der r-Wert rund 0.544 beträgt und somit ein schwacher direkter linearer Zusammenhang unter den Daten besteht.

Note: Auch hier nehmen wir wieder an, dass unser Ping 45 beträgt. Wir gehen nach dem gleichen Schema wie oben vor.

```
Switch to X[y] ;
```

Input ▹

```
Clear[x];
regrLineY[x_] = Fit[data, {1, x}, x];
regrLineY[45]  (* x value MUST be in data x-range *)
```

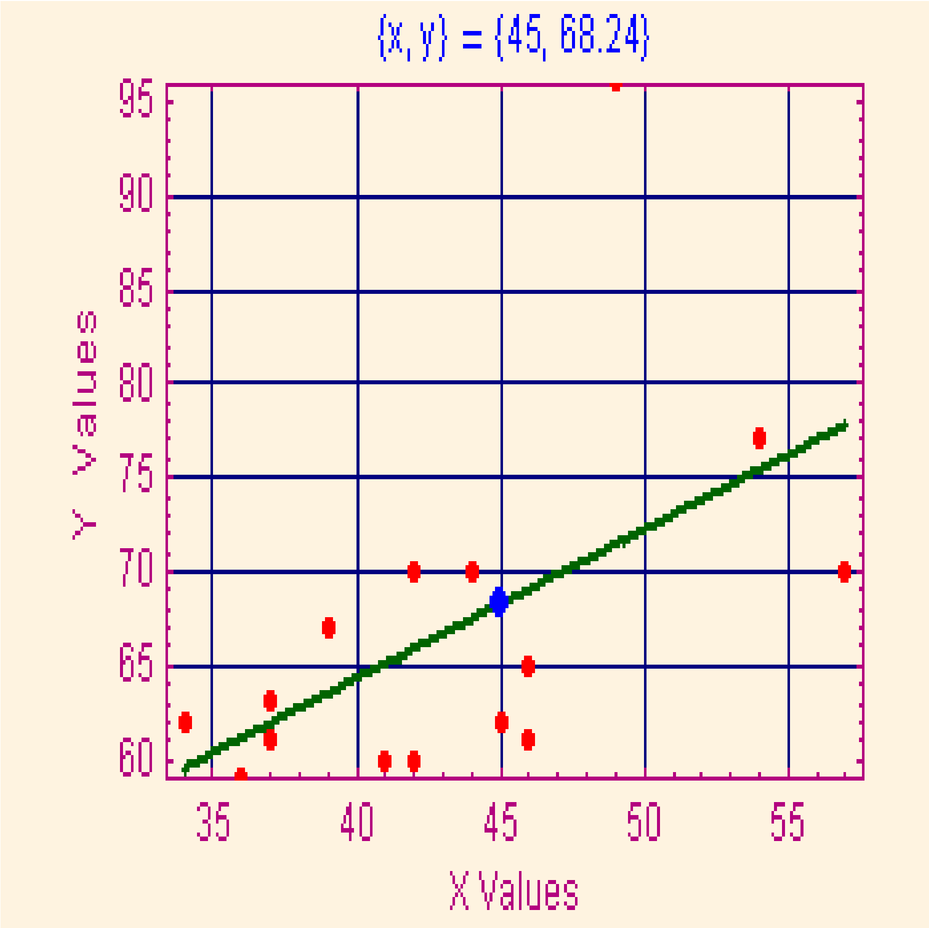
```
68.24
```

```
Switch to X[y] ;
```

Input ▹

```
xvalue = 45; (* xvalue MUST be in data x-range *)

MDSPlotDataRegressionLineYPt[data, xvalue,
    Epilog → {Blue, PointSize[.03]}];
```



Answer: Unser Ping von 45 entspricht, verglichen mit dem Mitspieler, der durchschnittlich einen Ping von 25 mehr als wir hat, 68.24.

Summary

Open / Close

Wir haben also zweimal den Ping von Mitspielern mit dem unseren verglichen.

Bei dem Ping der dem unseren ähnelte bestand ein starker direkter linearer Zusammenhang. Bei dem Ping der um ca 20-25 Einheiten höher war als unserer besteht ein schwacher direkter Zusammenhang. Dies ist darauf zurückzufahren, dass sein Ping mehr schwankt als der Ping eines Mitspielers, der den ähnlichen hat wird.

Im Großen und Ganzen kann man sagen, dass wenn es keine unerwarteten Störungen der Verbindungen gibt, die Pings sich ca. im gleichen Ausmaß verändern werden.

3 Reallife-Beispiel zur Normalität von Daten

Open / Close Print

Aufgabenstellung

Open / Close

Als begeisterte Computerspieler interessiert uns schon seit längerem, wovon der Ping abhängig ist. Auf Server zu spielen, bei denen der Ping zu hoch ist - oder der Ping so dermaßen schwankt, dass ein vernünftiges spielen unmöglich ist, ist keine Freude.

Aus diesem Grunde haben wir uns die ganze Sache einmal ein bisschen näher angeschaut.

Da wir im Rahmen unseres Mathematik Unterrichts die Möglichkeit bekamen Daten zu analysieren, beschlossen wir möglichst viele - und möglichst unterschiedliche Varianten des Pings zu notieren und auszuwerten.

Hier stellen wir uns der Frage, ob die Normalität der Daten gegeben ist.

Um eben unterschiedliche Ping Varianten zu bekommen beschlossen wir zwei Server aufzusuchen die räumlich sehr weit voneinander und zusätzlich auch sehr weit von uns (Graz) getrennt sind.

Der erste Server ist ein deutscher Server der in Berlin steht. Wir nehmen an, dass wir zu diesem Server eine relative gute Verbindung haben. (dazu müsste der Ping relativ klein sein, auf jeden Fall müsste der Ping einen Wert unter 50 annehmen)

Als zweiten Test-Server enttschieden wir uns für einen mit einem Sitz iergendwo in Großbritannien. - Der genau Sitz lies sich nicht eruieren, ist aber für unseren Versuch auch nicht weiter von Bedeutung.
Zunächst hatten wir zwar als zweiten Test-Server ein amerikanisches Gerät geplant - jedoch lies sich aufgrund der zu hohen und absolut unregelmäßigen Ping Daten keine vernünftige Analyse durchführen.

Mathematische Modelierung

Open / Close

deutscher Server:

Note: Wir speicherten also unsere Daten unter der Variable „data“.Ein einziger Klick auf die Palette **„Assessing Normality of Data“** vereinfacht uns die Variable zu modifizieren.Nach dem eintippen der Daten sieht dies aus wie unten angeführt.

```
data = { 49, 42, 47, 43, 47, 50, 38, 49, 30,
  49, 43, 43, 44, 44, 51, 40, 39, 45, 40, 39, 39, 48,
  44, 40, 45, 40, 38, 46, 44, 32, 33, 43, 36, 42, 44, 49, 44, 38, 36,
  47, 40, 45, 37, 49, 49, 40, 45, 40, 44, 40, 42, 47, 39, 40, 37, 33,
  48, 35, 38, 44, 42, 42, 46, 35, 35, 40, 37, 36, 42, 40, 38, 42, 46,
  45, 46, 49, 41, 49, 46, 48, 42, 44, 40, 39, 37, 50,
  46, 45, 44, 45, 42, 35, 40, 37, 46, 35, 39, 38, 50, 47,
  36, 38, 39, 37, 36, 39, 48, 35, 34, 36, 38, 40, 38,
  42, 42, 46, 40, 35, 35, 40, 37, 36, 42, 40, 38, 42, 46, 42, 39,
  34, 39, 38, 36, 38, 42, 36, 38, 43, 36, 39, 38, 32, 31, 41, 45,
  34, 47, 41, 37, 40, 39, 46, 50, 36, 33, 46, 33, 32, 40, 34, 40,
  38, 35, 40, 40, 43, 28, 39, 24, 40, 42, 44, 32, 34, 45, 40, 45,
  43, 43, 40, 40, 45, 43, 36, 43, 44, 32, 47, 43, 35, 38, 40, 43,
  47, 44, 48, 41, 42, 40, 45, 52, 57, 41, 45, 38, 48, 40, 43, 51,
  31, 37, 45, 46, 38, 53, 39, 43, 32, 31, 42, 40, 34, 26, 40, 28,
  45, 47, 35, 48, 40, 36, 36, 54, 32, 57, 41, 39, 38, 24} ;  Add
```

Um eine handfeste Überprüfung der Daten zu gewährleisten, stehen uns drei Möglichkeiten zur Verfügung:

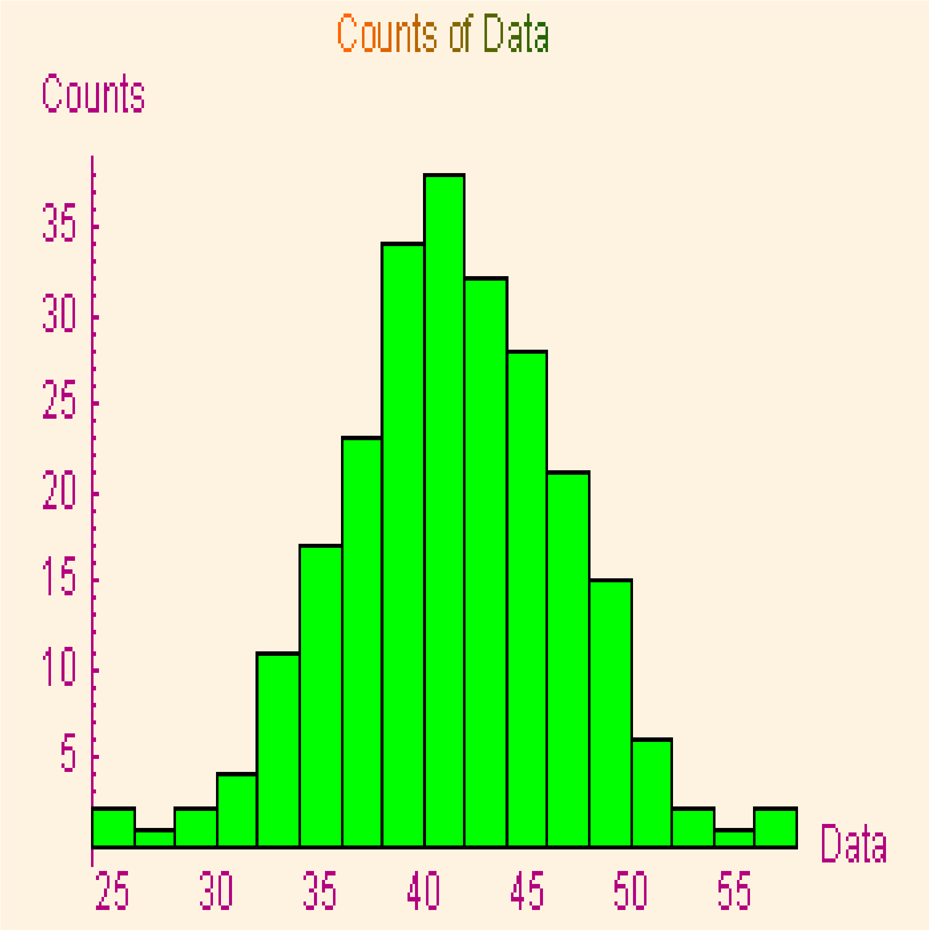
1. Histogramm

2. Sigma-Umgebungen

3. Normalwahrscheinlichkeitsplot

Note: Bei einem **Histogramm** werden alle Daten in Balken angezeigt. Je geringer die Datenunterschiede sind, desto genauer werden diese ausgegeben.

Input ▹
```
MDSHistogram[ data ] ;  (* data format like {3,-2.2,1/2,...} *)
```



Wie wir hier sehen können, wird auf der x-Achse der Wert des Pings ausgegeben. Die Anzahl, welcher Wert wie oft vorkommt, finden wir auf der y-Achse. Diese Grafik verdeutlicht uns, dass der Ping eine Normalität aufweist. Der Ping ist durchschnittlich zwischen 40 und 45 hoch. Natürlich treten auch Schwankungen auf. Diese Schwankungen sind auf dem deutschen Server jedoch sehr gering. Sie wirken sich nach unten und oben circa im gleichen Ausmaß auf das Histogramm aus.

Note: Unsere zweite Art der Prüfung auf Normalität ist die wie weit unsere Werte von der erwarteten **68%-95%-99,7% Regel** abweicht. In der 1. Sigma-Umgebung sollten sich 68% der Daten befinden, in der 2. Sigma-Umgebung 95% und in der 3. Sigma-Umgebung 99,7%. Die Berechnung wird vom Computer vorgenommen. Kleiner Abweichungen bedeuten aber nicht, dass es sich um keine Normalverteilung handelt.

Input ▹
```
MDSCheckStandardDeviation[ data ];
```

s – Interval Report

Interval	Percentage of Measurement	Expected
$\bar{x} \pm s$	69.8745 %	68 %
$\bar{x} \pm 2s$	95.8159 %	95 %
$\bar{x} \pm 3s$	99.1632 %	99.7 %

Bei unseren Daten, wie man oben erkennen kann, werden die erwarten Wert fast erreicht. Es sind lediglich Abweichungen von 1,8745%, 0,8159% und 0,5368% feststellbar. Also zeigt uns auch die zweite Überprüfungsmethode, dass wir hier von einer Normalverteilung sprechen können.

Note: Um zu erfahren, welchen Ping-Bereich die 1. Sigma-Umgebung umfasst, müssen wir zuerst μ mit $\bar{x}$ und σ mit s simulieren.

Input ▹

```
x̄ = Mean[data]; s = StandardDeviation[data];
n = Length[data];

MDSElementaryStatistics[ data];
```

Elementary Statistics

n	$\bar{x}$	s	SE Mean $s/\sqrt{n}$	s^2
239	40.7992	5.50109	0.355836	30.262

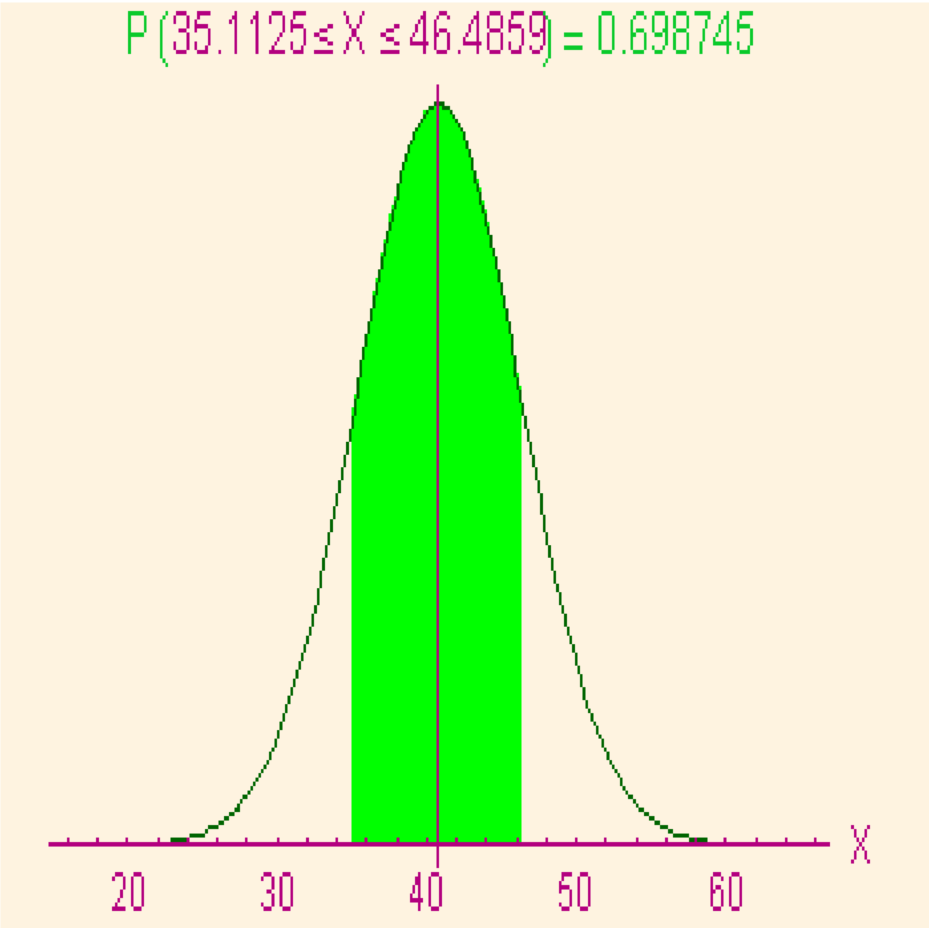
Answer: Unser $\bar{x}$-Wert beträgt 40.7992 und unser s-Wert ist 5.50109. Also umfassen die Sigma-Umgebungen auf den Ping bezogen die jeweilig unten angegebenen Werte:
1. Sigma-Umgebung von **35.2981** bis **46.3003**
2. Sigma-Umgebung von **29.797** bis **51.8014**
3. Sigma-Umgebung von **24.2959** bis **57.3025**

Note: Da wir wissen, dass die 1. Sigma-Umgebung 69,8745% aller Daten umfasst, können wir diese in eimem Plot grafisch darstellen.

```
Switch;
μ = 40.7992; σ = 5.50109;
probability = 0.698745;
Clear[z]; z =
  z /. FindRoot[2 MDSϕ[μ + z σ, {μ, σ}] - 1 == probability, {z, 0}];

MDSNormalDistributionPμσ[ μ - z σ ≤ X ≤ μ + z σ, {μ, σ}]
(* P (μ - ? ≤ X ≤ μ + ? ) = probab *)
```

Input ▹



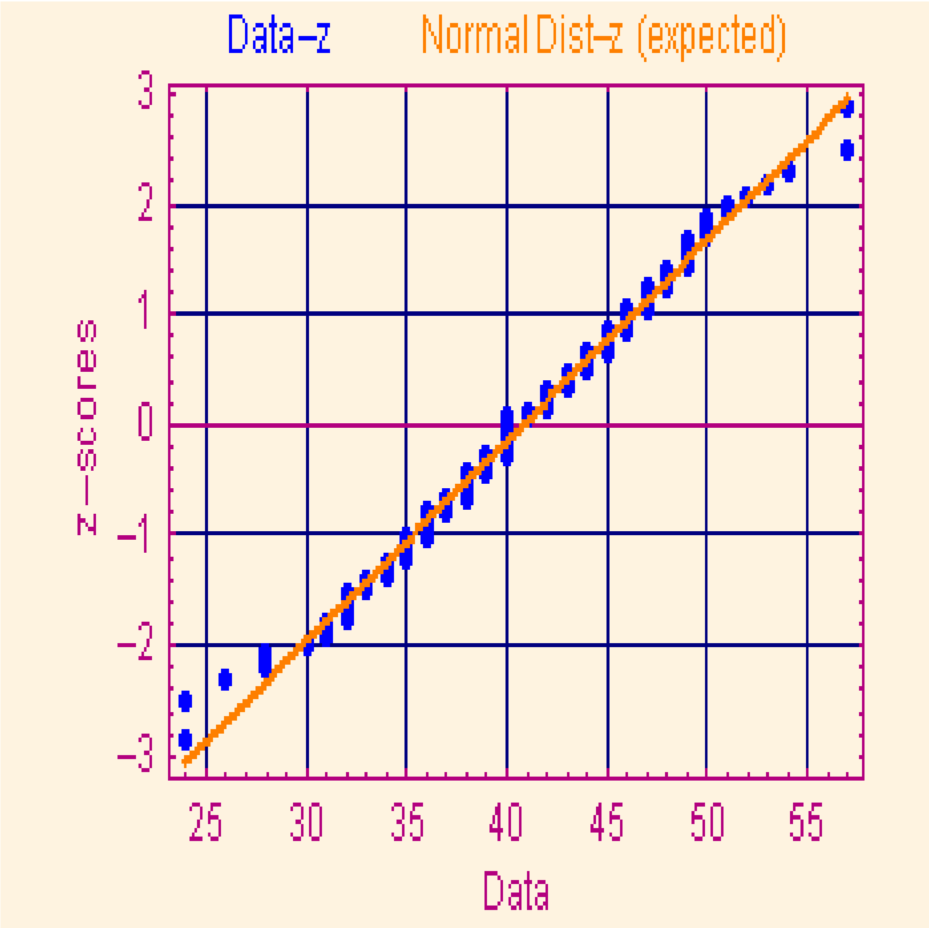
Normal Distribution

z (35.1125)	z (46.4859)	P (35.1125 ≤ X ≤ 46.4859)
-1.03375	1.03375	0.698745
Dev (z σ)		
40.7992 ± 5.68673		

Answer: Der grün markierte Teil der **Gaußschen Glockenkurve** spiegelt unsere 1. Sigma-Umgebung wieder. Diese hat einen z-Wert von -1.03375 bis 1.03375.

Note: Bei einem **Normalwahrscheinlichkeitsplot** wird eine erwartete Trendgerade durch ein Diagramm gezogen. Diese Trendgerade kann man mit Hilfe von Intervallen und den z-Werten berechnet werden. In sehr verständlicher Form ist uns dies im Unterricht beigebracht worden. Die Datenpunkte werden auf dem gleichen Diagramm eingezeichnet auf Grund der x-Achse, auf welcher unsere Werte eingezeichnet sind und der y-Achse, die den z-Wert angibt. Wenn die eingezeichneten Punkte nicht all zu weit von der Trendgerade entfernt sind, spricht man hier von einer Normalverteilung.

Input ▹ `MDSNormalProbabilityPlot[ data ];`



Nach Zeichnung dieses Plot sehen wir, dass die Werte des Pings stark bei der erwarten Trendgerade liegen. Wie erwartet sind ein paar Ausreißer erkennbar. Diese darf man aber bei der Beurteilung, ob es sich um eine Normalverteilung handelt, nicht zu stark beachten.

englisher Server:

Note: Bei einem englischen Server erwarten wir uns durch die weitere Entfernung unseres Computers zum Server eine größere Schwankung des Pings. Zu Beginn der

Auswertung gaben wir die erhaltenen Werte des englischen Servers in unser Notebook ein. Die Zahlen die durch den „lag“ entstanden sind, sind rot markiert.

Input ▹

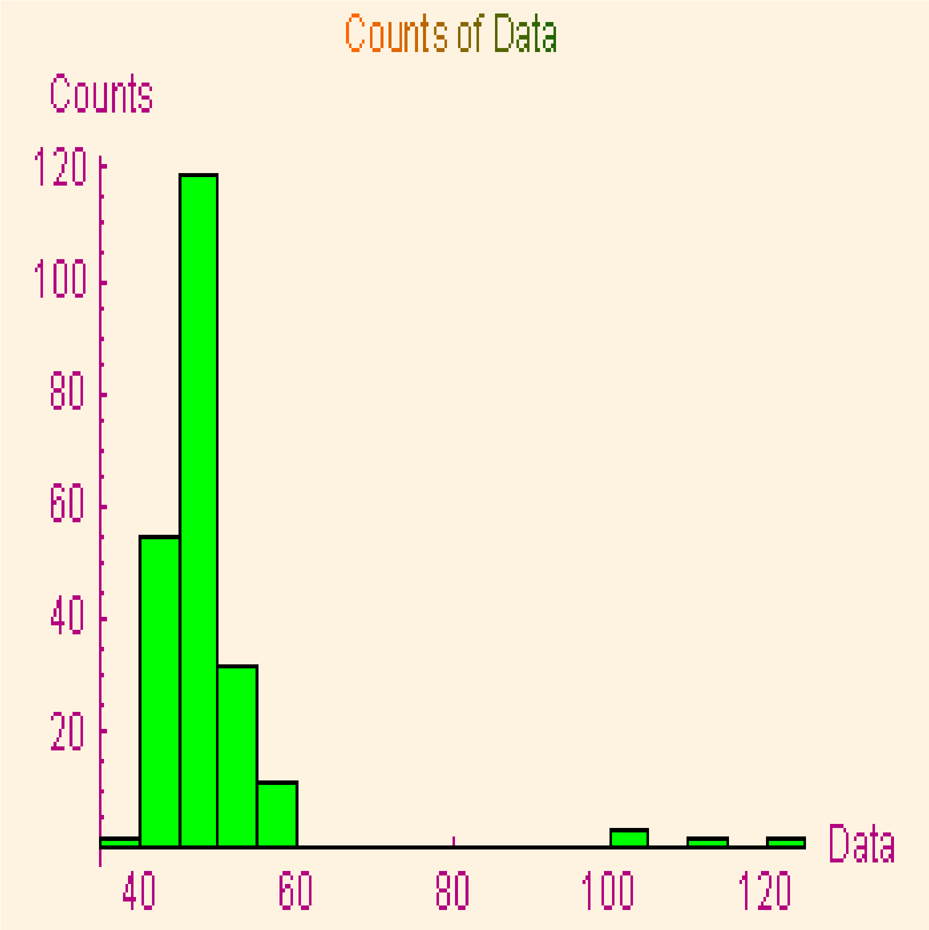
```
data = {56, 51, 48, 45, 49, 47, 48, 51, 45, 42, 46, 52, 46, 50, 46,
  45, 48, 44, 49, 43, 45, 47, 49, 42, 44, 46, 42, 46, 48, 47, 45,
  44, 43, 46, 48, 44, 46, 48, 51, 44, 45, 48, 51, 42, 46, 44,
  43, 49, 43, 44, 45, 43, 48, 44, 45, 44, 43, 51, 49, 45, 50,
  45, 44, 50, 46, 47, 45, 44, 45, 44, 49, 44, 56, 45, 45, 48,
  43, 51, 48, 42, 46, 48, 46, 47, 46, 47, 44, 38, 48, 47, 48,
  44, 46, 45, 58, 45, 50, 48, 50, 46, 48, 50, 46, 52, 54, 47,
  44, 46, 51, 44, 46, 53, 43, 49, 45, 43, 49, 45, 43, 57, 49,
  45, 46, 55, 44, 51, 46, 42, 43, 48, 46, 48, 45, 44, 42, 46,
  47, 50, 47, 43, 46, 47, 46, 56, 51, 48, 45, 49, 47, 48, 51,
  45, 42, 46, 52, 114, 102, 123, 100, 102, 46, 57, 48, 51, 42,
  43, 46, 44, 43, 49, 43, 45, 43, 48, 44, 46, 45, 58, 45, 50,
  48, 50, 46, 48, 50, 46, 52, 54, 47, 44, 46, 51, 44, 46, 53,
  43, 49, 45, 43, 57, 49, 45, 46, 55, 56, 51, 46, 42, 43, 48,
  46, 48, 45, 44, 42, 45, 47, 50, 47, 43, 46, 47, 46}; Add ;
```

Auch hier verwenden wir die gleichen Methoden, um dies festzustellen.

Note: Und hier werten wir die ermittelten Daten unter Zuhilfenahme eines **Histogrammes** aus.

Input ▹

```
MDSHistogram[data]; (* data format like {3,-2.2,1/2,...} *)
```



Das Histogramm weißt eine Normalität der Daten auf. Da die höheren Datensätze auch eingezeichnet werden, sind die geringeren Pingwerte in Balken, die fünf verschiedene Daten zusammenfassen, gegliedert.

Note:Mit Hilfe der **68%-95%-99,7% Regel** besteht für uns die Möglichkeit zu überprüfen, ob es sich bei den erruirten Daten um eine Normalverteilung handelt.

Input ▹
```
MDSCheckStandardDeviation[ data ];
```

s – Interval Report

Interval	Percentage of Measurement	Expected
$\bar{x} \pm s$	97.3094 %	68 %
$\bar{x} \pm 2s$	97.7578 %	95 %
$\bar{x} \pm 3s$	97.7578 %	99.7 %

Durch den vorgekommen „lag“, der die Pingwerte auf über 100 ankurbelte berechnet der Computer die erwarteten Sigma-Umgebungen vollkommen anders. Deshalb fallen nun fast alle Daten in die 1. Sigma-Umgebung. Die 2. Sigma-Umgebung ist der 3. gleich.

Note: Mit Hilfe des $\bar{x}$-Wertes und des s-Wertes können wir die Sigma-Umgebungen berechnen.

Input ▹
```
x̄ = Mean[data]; s = StandardDeviation[data];
n = Length[data];

MDSElementaryStatistics[ data];
```

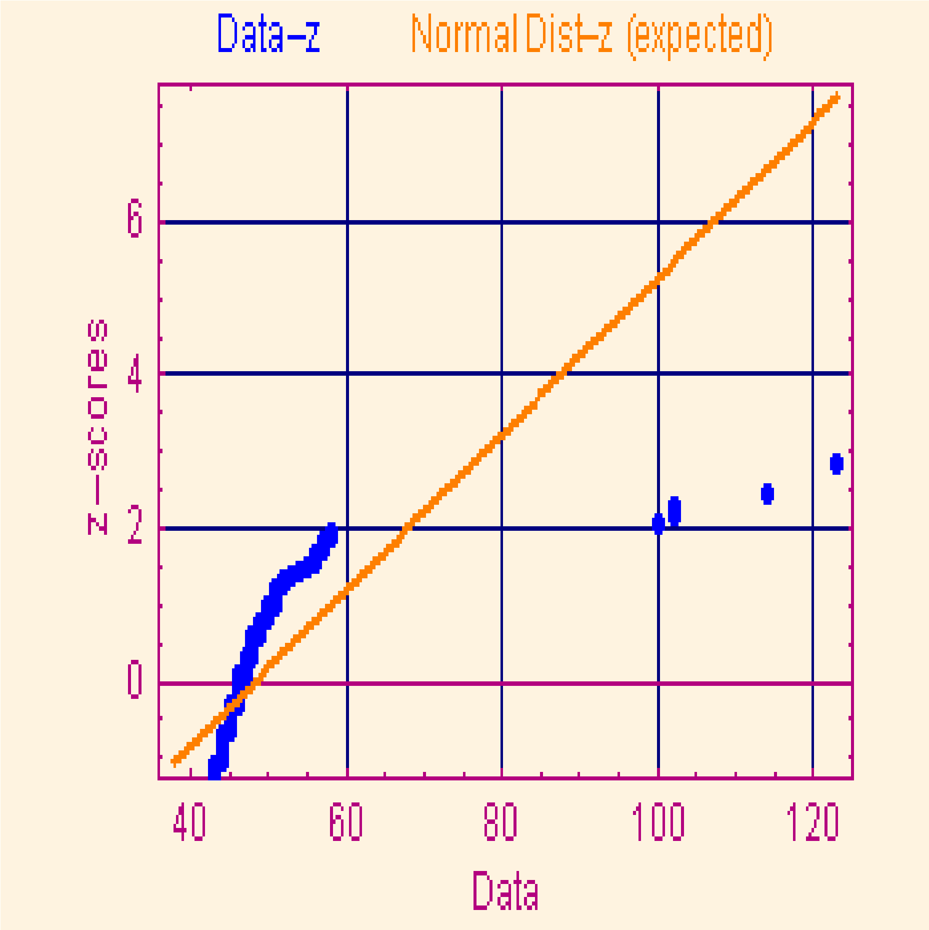
Elementary Statistics

n	$\bar{x}$	s	SE Mean $s/\sqrt{n}$	s^2
223	48.2646	9.83178	0.658385	96.6639

Answer: Unser $\bar{x}$-Wert beträgt 40.7992 und unser s-Wert ist 5.50109. Also umfassen die Sigma-Umgebungen auf den Ping bezogen die jeweilig unten angegebenen Werte:
1. Sigma-Umgebung: **48.2646 ± 1 * 9.83178**
2. Sigma-Umgebung: **48.2646 ± 2 * 9.83178**
3. Sigma-Umgebung: **48.2646 ± 3 * 9.83178**

Note: Als dritte und zugleich auch letzte Überprüfungsvariante haben wir den **Normalwahrscheinlichkeitsplot** zur Verfügung. Hierbei lässt sich ungefähr mit dem Auge feststellen, ob die ermittelten Daten normal verteilt sind beziehungsweise wie die Daten verteilt sein hätten müssen um normal verteilt zu sein.

Input ▹
```
MDSNormalProbabilityPlot[ data ];
```



Der Normalwahrscheinlichkeitsplot zeigt uns deutliche Abweichungen von der erwarteten Trendgerade. Die Punkte von 47 bis 58 zeigen einen zu hohen z-Wert auf. Alle anderen Punkte liegen unter dem erwarteten z-Wert.

Summary

Open / Close

deutscher Server:

Nach Durchsicht und genauer Studierung der verwendeten Varianten zur Beurteilung der Daten, sind wir auf den klaren Beschluss gekommen, dass es sich um eine Normalverteilung handelt. Alle drei Methoden der Ermittlung haben uns dies bestätigt. Der Ping hat einen bestimmten Mittelwert der nach oben und unten im circa den gleichen Ausmaß schwankt. Ohne das Auftreten eines "lags" kommt keine große Abweichung vor.

englischer Server:

Auf den ersten Blick sehen die Daten auf keinen Fall normal verteilt aus. Weder erwarteten Sigma-Umgebungen werden erreicht, noch liegen die eingezeichneten Punkte in der Nähe der erwarteten Trendgerade. Einzig das Histogramm lässt noch den möglichen Schluss auf eine Normalverteilung ziehen. Also wurden wir nur aufgrund der kurzzeitigen Schwankung - dessen Ursache sich unmöglich eruieren lässt - unserer Hoffnung auf eine Normalverteilung beraubt.

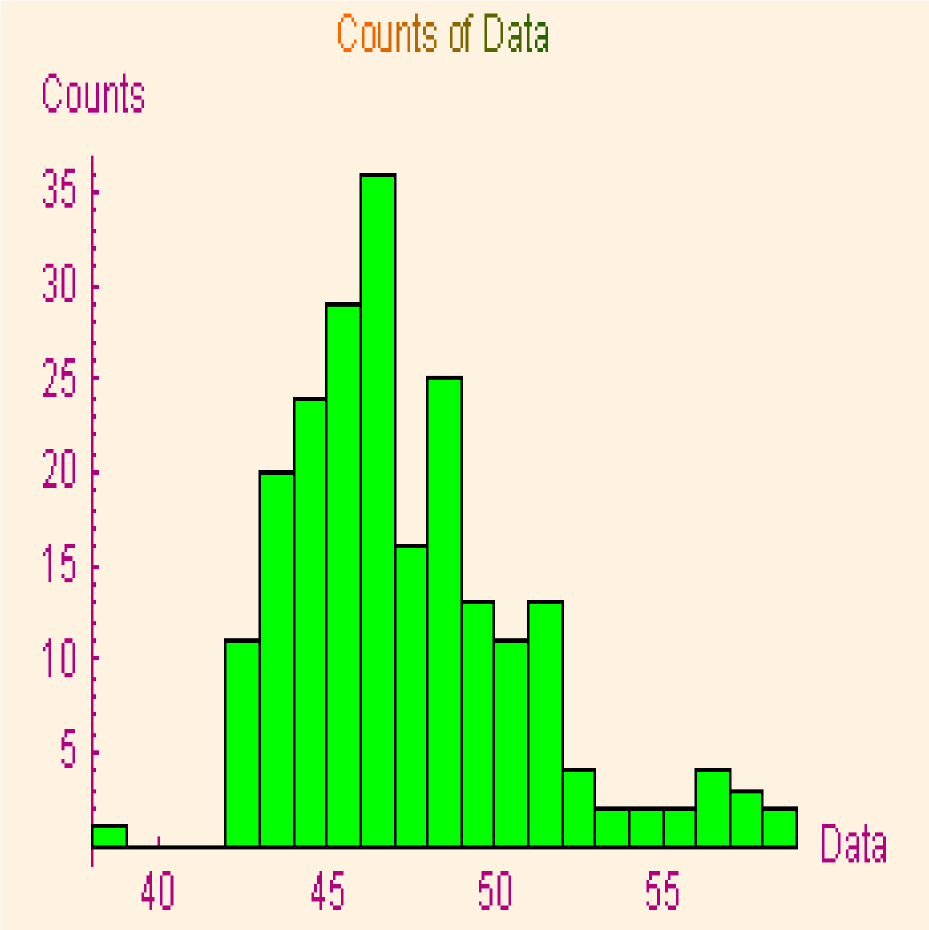
Aufgrund dieser Erkenntnis haben wir einen kleinen Test durchgeführt. Bei diesem haben wir den kurzzeitigen „lag“ nicht bei der Berechnung der drei Methoden einfließen lassen. Kaum waren diese hohen Werte entfernt, sahen unsere Daten normal verteilt aus.

Note: Jetzt zeigen wir die Daten des englischen Servers nach Entfernung des "lags" an.

Input ▹
```
data = { 56, 51, 48, 45, 49, 47, 48, 51, 45, 42, 46, 52, 46, 50, 46,
  45, 48, 44, 49, 43, 45, 47, 49, 42, 44, 46, 42, 46, 48, 47, 45,
  44, 43, 46, 48, 44, 46, 48, 51, 44, 45, 48, 51, 42, 46, 44, 43,
  49, 43, 44, 45, 43, 48, 44, 45, 44, 43, 51, 49, 45, 50, 45, 44,
  50, 46, 47, 45, 44, 45, 44, 49, 44, 56, 45, 45, 48, 43, 51, 48,
  42, 46, 48, 46, 47, 46, 47, 44, 38, 48, 47, 48, 44, 46, 45, 58,
  45, 50, 48, 50, 46, 48, 50, 46, 52, 54, 47, 44, 46, 51, 44, 46,
  53, 43, 49, 45, 43, 49, 45, 43, 57, 49, 45, 46, 55, 44, 51, 46,
  42, 43, 48, 46, 48, 45, 44, 42, 46, 47, 50, 47, 43, 46, 47, 46,
  56, 51, 48, 45, 49, 47, 48, 51, 45, 42, 46, 52, 46, 57, 48, 51,
  42, 43, 46, 44, 43, 49, 43, 45, 43, 48, 44, 46, 45, 58, 45, 50,
  48, 50, 46, 48, 50, 46, 52, 54, 47, 44, 46, 51, 44, 46, 53,
  43, 49, 45, 43, 57, 49, 45, 46, 55, 56, 51, 46, 42, 43, 48,
  46, 48, 45, 44, 42, 45, 47, 50, 47, 43, 46, 47, 46} ;  Add ;
```

Histogramm

Input ▹
```
MDSHistogram[ data ] ;  (* data format like {3,-2.2,1/2,...} *)
```



Sigma-Umgebungen

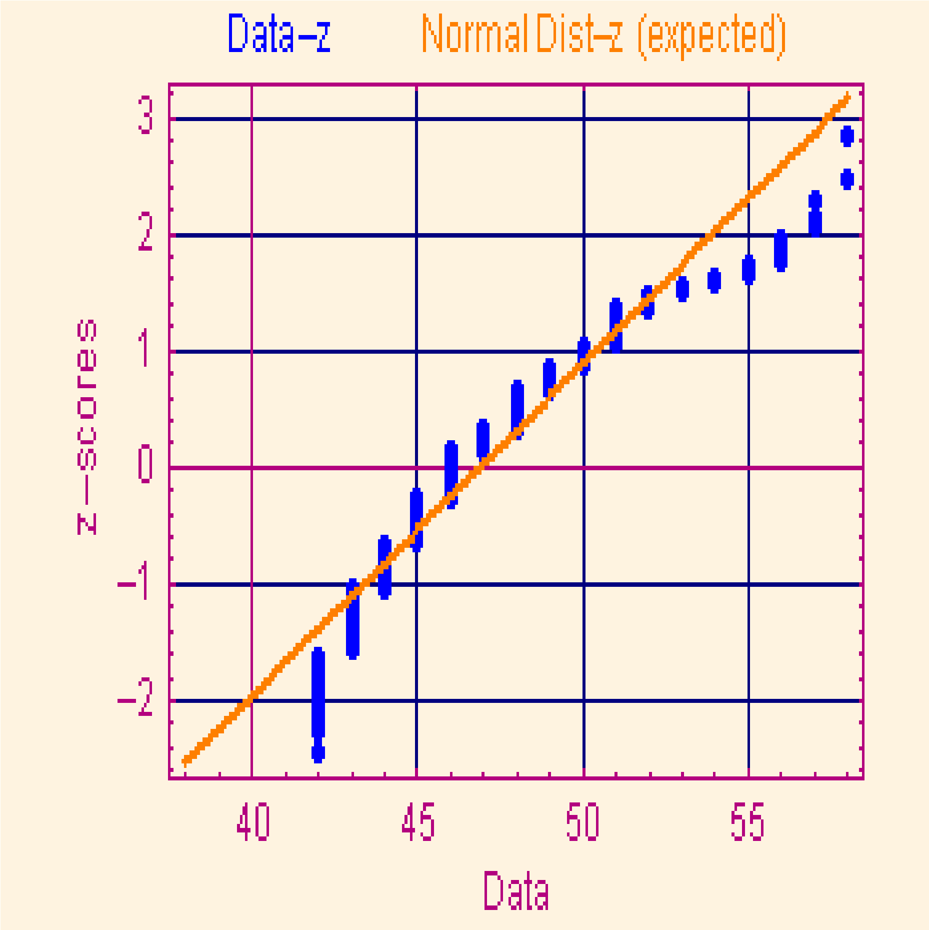
Input ▹ `MDSCheckStandardDeviation[ data ];`

s – Interval Report

Interval	Percentage of Measurement	Expected
$\bar{x} \pm s$	70.6422 %	68 %
$\bar{x} \pm 2\,s$	93.578 %	95 %
$\bar{x} \pm 3\,s$	99.0826 %	99.7 %

Normalwahrscheinlichkeitsplot

Input ▹ `MDSNormalProbabilityPlot[ data ];`



New Chapter Cut Last Chapter